DOCUMENT RESUME

ED 158 844 PS 009 547

AUTHOR Miley, Alan
TITLE Procedures for Classroom Observations: 1973-1974. Technical Report #19.
INSTITUTION Kamehameha Schools, Honolulu, Hawaii. Kamehameha Early Education Project.
SPONS AGENCY California Univ., Los Angeles. Mental Retardation Research Center.; Hawaii State Dept. of Education, Honolulu.
PUB DATE [74]
NOTE 41p.; Tables are filmed from best available copy; For related documents, see PS 009 533-546 and PS 009 548-573

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Behavior Patterns; Behavior Rating Scales; *Classroom Observation Techniques; Contingency Management; *Demonstration Programs; Early Childhood Education; *Grade 1; Hawaiians; Interaction Process Analysis; *Kindergarten Children; *Student Behavior; *Teacher Behavior
IDENTIFIERS Hawaii; *Kamehameha Early Education Program

ABSTRACT

This report describes classroom observation techniques used to record the behavior of educational specialists (teachers) and students in a kindergarten and a first grade classroom of the Kamehameha Early Education Program (KEEP). Classroom behavior was observed and recorded daily during the 1973-1974 school year. Each student was observed three times during each of three daily observation sessions. Behavior during these observation sessions was designated as on-task, off-task, or disruptive. Three coding systems were used successively during the year for recording teacher behavior. Difficulties encountered in the first coding system led to the modifications used in systems II and III. Each of these coding systems focused on the teacher's reinforcement of student behavior. The reliability of the teacher and student coding systems was analyzed by comparing the coding records of independent observers. (BD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED158844

Technical Reports

of

The Kamehameha Early Education Program

a research and development program established and funded by

The Kamehameha Schools/Bernice P. Bishop Estate

Ronald Gallimore, Roland G. Tharp & Gisela E. Speidel,
General Editors

Ellen Antill
Production Editor

Technical Report #19

The cooperation of the State of Hawaii Department of Education is gratefully acknowledged, as is the support and resources made available by the Sociobehavioral Research Group, MRRC, University of California, Los Angeles.

PS 009547

The opinions expressed herein do not necessarily reflect the position, policy or have the endorsement of The Kamehameha Schools/Bernice P. Bishop Estate, or of the editors.

Published by The Kamehameha Early Education Project, 1850 Makuakane Street, Honolulu, HI 96817

All rights reserved. No parts of this report may be reproduced in any form or by any means without the prior written permission of The Kamehameha Schools/Bernice P. Bishop Estate.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Kim C. M. Sloat

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The Kamehameha Early Education Program

The Kamehameha Early Education Program (KEEP) is a research and development program of The Kamehameha Schools/Bernice P. Bishop Estate. The mission of KEEP is the development, demonstration, and dissemination of methods for improving the education of Hawaiian and Part-Hawaiian children. These activities are conducted at the Ka Na'i Pono Research and Demonstration School, and in public classrooms in cooperation with the State Department of Education. KEEP projects and activities involve many aspects of the educational process, including teacher training, curriculum development, and child motivation, language, and cognition. More detailed descriptions of KEEP's history and operations are presented in Technical Reports #1-4.

Abstract

This report describes classroom observation undertaken at KEEP during the school year 1973-1974. Behavior of both students and educational specialists (ES) who functioned as teachers, was observed and recorded daily from September 24, 1973, until June 7, 1974. Student observations focused on task orientation and disruptive behavior. ES observation focused on reinforcement behavior.

The report includes a general description of the observation situation, a description of the codes used, including categories and method of observation, an analysis of some unsuccessful attempts in teacher code development, and a discussion of reliability.

Technical Report #19

Procedures For Classroom Observations: 1973-1974

Alan Miley

This report describes classroom observation undertaken at KEEP during the school year 1973-1974. Behavior of both students and educational specialists (ES) who functioned as teachers was observed and recorded daily from September 24, 1973, until June 7, 1974. Student observations focused on task orientation and disruptive behavior. ES observation focused on reinforcement behavior.

This report includes a general description of the observation situation, a description of the codes used, including categories and method of observation, an analysis of some unsuccessful attempts in teacher code development, and a discussion of reliability.

Subjects

The behavior of six educational specialists who served as teachers was monitored during these observations. Student population varied throughout the year, with several students withdrawing and others entering in mid-year. The size of the kindergarten class ranged from 26 to 27, half of whom were male and half were female. The size of the first grade class ranged from 24 to 28 children, at its largest, 12 male and 16 female.

Observers

Observers of student behavior included, EA, KB, LK, and SO with the majority of observations conducted by EA and LK. Observers of teacher behavior included EA, LK, and AM. A final observer, LW, conducted observations of both student and teacher behavior at other schools.

Setting

Two classrooms, comparable to those in DOE schools except for observation facilities, were the site of the observations. Each classroom was equipped with suspended overhead microphones and remote controlled videotape cameras which channelled sound and images into an elevated observation theatre. In addition, one-way glass, high on the classroom walls, permitted views of both classrooms from the observation theatre.

All student behavior observations were made through the one-way glass from the observation theatre. Most observations of teacher behavior were also conducted from the theatre. Sound was most commonly transmitted by a wireless microphone worn by the teacher. Occasionally, depending on availability and performance of sound equipment, mobility and voice volume of the teacher, and competition for observation facilities, other methods of observation were used. Such methods included, in order of decreasing use, observation through the one-way glass with sound from the suspended microphone, in-classroom observation, and observation "live" on a TV monitor with sound from the suspended microphones.

Scheduling

Several different observation schedules were in effect during the course of the year. Changes in schedule were a result of changes of daily teaching schedule, observer availability, and of experimental classroom procedure.

To obtain a representative sample of the different classroom settings, observation schedules took account of such variables as day of the week, time of the day, subject matter, classroom configuration (small group, learning center, large group, etc), teacher-student interactions, and coincidence vs. independence. Resulting schedules were cycles which were repeated every one to four weeks.

In general, student behavior was observed three sessions per day during various subjects, with various teachers, etc. Observations of teacher behavior were limited to the two new teachers from 10-4-73 to 2-1-74. Each of these teachers was usually observed once on each of three days per week. Commencing on 2-4-74, each of the six teachers was observed daily. Again, the schedule was designed to vary subject matter, student group, etc., as described above.

Student Behavior Observation Categories

The following mutually exclusive response classes of student behavior were of interest:

(a) On-Task (coded +)

(b) Off-Task (coded -)

(c) Disruptive (coded D)

Together these categories were designed to encompass all pupil behaviors which may occur in the classroom. Explanations of each are:

On-task -- any behavior appropriate to the task demands of the particular situation. This mainly involved working on an assignment, attending to the teacher, following instructions, and cooperative work with peers (i.e. talking with peers concerning the task at hand). Such behaviors as nondisruptive travelling from one table to another, pencil sharpening, etc. were coded as "on-task" when these behaviors were appropriate to the task demands of the particular classroom situation. Play behaviors were considered on-task when appropriate, i.e. when it was approved by the teacher following completion of work, and it was not disruptive (see below).

Off-Task -- any behavior inappropriate to the task demands of the particular situation (excepting "disruptive" [see below]). This mainly involved daydreaming, watching others, looking around the room, or being physically away from the designated work area.

Disruptive -- this is a special form of "off-task" involving behaviors which interfere with either another child or the teacher or both. In general, this included shouting, crying, fighting, running across the room, and inappropriate socializing (i.e. talking to another child about matters unrelated to their tasks).

Student Behavior Observations: Order of Observations

Rosters were prepared, with students listed according to their groups in the classroom.

Roster 1, in effect through December of 1973, consisted of two lists, one for kindergarten and one for first grade, with five groups on each list. Each day, one list was randomly chosen to begin the student observations of that day. Furthermore, on Mondays, a student observation cycle began with the first student in group 1 of the chosen list. Students were then observed sequentially on that list until the last student in group 5 was observed. Next, the first student in group 1 of the other list was observed and other students observed sequentially, until the last student in group 5 was observed and the cycle was complete. On Tuesdays, the cycle began with the first student in group 2 of the chosen list. Students were observed sequentially through group 5, then group 1. Next, the first student in group 2 of the other list was observed, etc. The cycle was complete when the last student in group 1 of this list was observed. On Wednesdays, group 3 was observed first, on Thursdays, group 4 was observed first, and on Fridays, group 5 was observed first.

On January 7, 1974, rosters 2 and 3 were introduced to reflect changes in classroom procedure. Each contained both kindergarten and first grade students. Roster 2, consisting of six groups of students, was used in morning sessions, and roster 3, consisting of three groups of students, was used in afternoon sessions.

For rosters 2 and 3, an observation cycle would start with group 1 on one day, group 2 on the next, etc. After group 6 for roster 2 and group 3 for roster 3, a new cycle would start with group 1 on the next day. Occasionally, these observation orders were disrupted (see Appendix A).

Though observation schedules were in effect, in actuality, the start of

the observation depended upon the start of classroom activities. For example, spring math was taught with three classes of students, meeting at the same time. Observations did not begin until all three classes had started. Occasional problems arose because of this procedure, especially during the spring learning center period when the two classes were merged. Students rotated in and out of reading groups and the two reading teachers were often not synchronized, one beginning a new group while the other was still with a previously scheduled one (see Appendix A). Attempts to coordinate student observations with observations of teacher behavior exacerbated the problem.

Student Behavior Observations: Method

When the conditions for beginning an observation session were met, the observer spotted the student to be observed first. Upon spotting the student, the observer took a "mental photograph" of his behavior at that instant. The observer then recorded a "+" if he was on-task, a "-" if he was off-task, or a "D" if he was disruptive, next to his name on the roster at the instant of the "photograph." If the student was not in the observer's field of vision (i.e. not in either classroom), no mark was recorded. The observer then spotted the next student on the roster and repeated the procedure.

After an entire cycle, with each student observed once, the cycle was repeated. An observation session comprised three cycles, i.e. each student was observed three times per session. In such a case, no more observations were made on those students who switched during that session.

The amount of time required for an observation session depended on the experience of the observer, whether the students were clustered in a large group or spread out in tables, the number of students present, etc. The average time was approximately 10 minutes and ranged from around five to fifteen minutes.

Teacher Behavior Observations: Unsuccessful Teacher Codes

Several codes of teacher behavior were attempted. The codes (discussed below) used for actual data collection focused on various methods of positive and negative consequences used by the teacher. Unsuccessful attempts were made at coding the appropriateness of reinforcement and instruction behavior such as modelling, asking questions, etc.

The attempt at assessing appropriateness was based on the idea that a reinforcement was inappropriate if:

1. Positive reinforcement was given to a student at a time when he was off-task or disruptive.
2. Punishment was given to a student when he was on-task.
3. In the case that reinforcement was delivered to a group of students (e.g. a table, an entire class), the behavior of the majority of the group was used to assess appropriateness.

One problem was that of discriminating a reinforcement vs. a prompt to action. If, for example, a student who has been sitting at his table working, begins wandering around the room. A teacher may say, "You know, I like the way you have been working this morning." Often this functions as a prompt to him to return to his table and work. At other times, this serves to reinforce his wandering behavior. The different functions depend on other factors, for example, the amount of positive reinforcement that he receives while he is working. This issue might have been resolved by operationalization if it had not been for the thornier problem mentioned below.

The major difficulty was the post hoc nature of assessment. The observers, focusing on teacher behavior, could pay little attention to the behavior of the students in the class, especially those out of the visual field which included the teacher. Outside of an occasional intuition, the observers

could not tell when a teacher was going to deliver a reinforcement until she began to do so. But, once a teacher began reinforcing a student, she usually gained his attention and off-task behavior, if any, terminated.

A possible method of overcoming this problem is use of a system of interactive observations, one where the attention is neither on the teacher nor the student, *per se*, but on their interaction. Such a system might involve watching each student for 30 seconds, observing his interactions with his total environment: his present task, his peers, and his teacher. From the point of view of teacher behavior, the result would be a time sampling. Her behavior would be noted only when she was interacting with the student in question. With the present system, however, the difficulty was insurmountable.

Another unsuccessful attempt was made to include teacher behaviors unrelated to reinforcement in the code. Several categories were attempted in the process, but the following four--informing, instructing, questioning, and modelling--demonstrate the basic approach.

Basically, informing is presenting information. An example of informing is the delivery of a lecture in a college. Instructing, in contrast to informing, requires a behavioral response from the student; it is a prompt to action. Questioning is an inquiry directed to a student. Modelling is the demonstration to a student of how a required response is to be done. Within each of these four categories, attempts were made to discriminate between an academic and a classroom management focus.

Discrimination between these mutually exclusive categories was quite difficult at first, and even with experience and a long list of conventions, remained difficult. The discrimination of instructing from informing was one of the more difficult ir this regard.

A second problem was encountered in quantifying units for these categories.

When, for example, a teacher reads the class a story, how many "informings" does she do? Difficulties in determining units arose much more often with these types of behaviors than with the reinforcement behavior which ultimately constituted the teacher code.

A third difficulty compounded these problems: the extremely high frequency of occurrence of behaviors to be coded. The "instructional" categories were to be coded simultaneously with the reinforcement behaviors described in teacher codes 1 and 2 below. The result was that a large proportion of the teacher's classroom behavior was to be recorded. In actual observation, some unit of behavior was recorded every three-four seconds. An observer, looking down at his paper to record a behavior, might miss the next behavior entirely. Frequency and difficulties in distribution and unit quantification resulted in extremely low reliabilities, seldom as high as .50, and the code was simplified to focus on reinforcement behaviors only.

Teacher Behavior Observations: Implemented Teacher Codes

During the 1973-1974 school year, three different teacher codes were used, designated teacher codes 1, 2, and 3. The codes were in effect on the following dates:

9-24-73 to 10-4-73	Teacher Code 1
10-5-73 to 2-6-74	Teacher Code 2
2-7-74 to 6-7-74	Teacher Code 3

Teacher Code 1

VA, VAs: Praise, supportive statements, and other verbal comments, implying positive evaluation related to academic work. VA refers to more general comments (Good job! You are a hard worker. Beautiful! What a good student!), while VAs refers to comments in which a specific behavior is noted (You're staying right on the lines. You answered that exactly right. Yes, that word begins with a "b").

VM, VMs: Praise, supportive statements, and other verbal comments implying positive evaluation related to classroom management. VM refers to more general comments (Kaliko knows how to get ready for class. The people at this table are behaving so well! I like your attitude), while VMs refers to comments in which a specific behavior is noted (I like the way you are sitting in your seat. Good, your eyes are to the front of the room. Keola, you're using such a nice quiet voice).

V-, V-s: Desists, scolds, and other verbal comments indicating negative evaluation or implying that a desired response has not occurred. V- refers to more general comments (My, you students are bad today. That's the worst work I've seen all year. Stop fooling around this instant!), while V-s refers to comments in which a specific behavior is noted (I can't continue while you're talking. Can't you even write an "A"?! Stop that hitting!)

NV: Nonverbal positive interactions including positive body contact (pats, hugs), and positive hand gestures (applause, A-OK sign). Privileges (being first in line, getting to collect the pencils) count as NV either when established as privileges in the setting or when the contingency is expressed verbally.

NV-: Nonverbal negative interactions including negative "desisting" body contact (slaps, physical restraint), negative hand gestures (clapping for quiet, hands to lips for quiet) and sending a student to time-out. Negative privileges count as NV- either when established as "nonprivileges" in the setting, or when the contingency is expressed verbally.

NVt: Tokens including stars on papers, tabs, candy, names written on board in list of good students.

NV-t: Negative tokens (taking tokens away, names written on the board in list of disruptive students).

Teacher Code 2

Ak: General verbal praise with positive tone acknowledging student's academic work (e.g. Okay! Perfect! Thank you! Beautiful! Good! Very good!) consisting of one or two word phrases.

Ag: General verbal praise about academic work, given in several words (e.g. You did good work today. I like the way you have been working. You have been working harder and harder.).

As: verbal praise specifying student's academic work (e.g. You are cutting well. You are writing well. Look at how straight a line he drew! [to class]).

MK: General verbal praise with positive tone acknowledging desirable nonacademic behaviors such as sitting, head on table, listening (when not in listening center), hands in lap, etc. consisting of one or two word

phrases (e.g. Good; Fine; Okay; etc. when referring to nonacademic behaviors. The decision to record behavior as academic or nonacademic depends greatly on the recorder's assessment of the situation.

Mg: General verbal praise about nonacademic behaviors using more than two words (e.g. You're behaving well. I like your attitude).

Ms: Verbal praise specifying student's nonacademic behavior (e.g. You're sitting very nicely. I like the way you are all looking at me. Everybody, look how well Keone is sitting!)

S: Scolds, as determined by tone (Think! How many times do I have to tell you what that number is?! I want that mess cleaned up this minute!)

SE: Scolds plus negative evaluation; a scolding tone in a statement that includes a negative evaluation of that child (e.g. You kids are really dumb! I don't know why I bother to teach you, you'll never learn!)

NE: Negative evaluation without the scolding tone (e.g. You're not doing that right; You're wrong; Class, look at how badly Jeremiah drew his picture.)

D: Desist statements, with or without the negative tone (Stop! No, don't color that picture. Ssh! I want this noise stopped this minute!)

P+: Granting privileges (e.g. giving the best student his choice of toy for recess--P+ recorded at the point at which teacher gives privileges; letting best children leave first). A privilege is assessed as one when its positive connotations are known to the class, either as an established precedent or when the contingency is expressed.

T+: Tokens (e.g. good work awards, stars on paper, writing names of good children on blackboard).

B+: Positive body contact (e.g. patting, shaking hands, hugging).

H+: Hand movement of teacher signifying her pleasure with child (e.g. clapping for child, A-OK sign).

P-: Negative privilege or loss of privilege.

T-: Negative tokens (taking tokens away, putting bad children's names on board).

B-: Negative body contact (guiding child's hand away from reaching for undesirable object; guiding child away from or toward area, with exception of to time-out; grabbing child's hand to stop him from hitting).

(B-): Painful negative body contact (hitting, slapping, pulling forcefully away so as to cause pain).

H-: Movement of teacher's hand signifying negative idea (e.g. placing

finger over mouth to signify quiet).

R: Time-out cue; any indication of time-out for child (e.g. pointing to time-out area; handing child time-out timer; bodily moving child to time-out area).

O: Taking away an inappropriate physical object from child.

In both teacher codes 1 and 2, important discriminations are those between general vs. specific praise (e.g. V- vs. V-s in teacher code 1). These discriminations, as well as rules for determining frequency, are explained below.

Academic vs. Management

Praise is considered management when it is contingent upon behaviors which are components of order (e.g. raising hands for asking questions) or which are conditions for enhancing learning (e.g. paying attention, eyes to the front of the room). Academic praise is contingent upon behaviors which are a significant part of the subject being taught. Praise for listening, for example, is considered academic when the student is at a listening center or in "story time," but it is considered management when instructions are being given, a math lesson is being taught, etc. In some cases (e.g. "My you are good today!"), the observer cannot tell from the comment per se whether an academic or management praise should be coded. The distinction must arise from his assessment of the general classroom situation ("reading the teacher's mind"). In cases where the observer cannot decide, academic is chosen over management.

Specific vs. General

When the behavior reinforced comprises several parts, it is general. When, on the other hand, the behavior is produced by a single action, it is specific. Most common "general" behaviors are working, trying hard, doing a good job, being a good student, and variants thereof. "Specific" behaviors, in contrast, include such things as listening, writing, answering correctly,

staying in the lines when coloring, etc.

In the case of teacher code 1, one or two word praises are considered general (e.g. "good"; "that's correct," etc.). Even in cases where the praise is for a specific piece of work or answer to a specific question, if the behavior is not specified by the teacher, it is considered general. Similarly, "you're wrong!" "That's bad," etc. are considered general.

Determining Frequency

1. When a teacher states different specific types of behavior, these behaviors are considered discrete (e.g. "I like the way that you are sitting and listening" counts as two praises). When a teacher states a general and specific type of behavior in juxtaposition, only the specific one is counted (e.g. "you have been so good today; you've stayed in your seat all morning" or "You wrote right on the line. That's good work!", each count as one specific praise).

2. When the teacher repeats a response, only one is counted, ("Good work, so good!") except for the following conditions:

   a. When she is pointing out or otherwise demonstrating that she is referring to different things (e.g. in going down a sheet of written work and praising each example).

   b. When she intersperses another comment in between (e.g. "Good work, Kalani;" "Keola, good work;" "Good work, Kalani;" this counts as three all together).

   c. When at least 10 seconds intervene between the responses.

3. A reinforcement given to a group of children, as a group, counts as a single response. If, on the other hand, group members are specified, each

specification counts as a response. For example, if Table 5 includes Ho'oulu, Momi, and Nani, "Table 5 is really working hard!" counts as a single praise, but "Ho'oulu, Momi, and Nani are really working hard" counts as three praises.

4. When a single behavior occurs for 10 seconds, it is counted as a new occurrence at the end of each 10 seconds.

Training VISTA volunteers for observations of teacher behavior in the Big Island was extremely difficult. The volunteers had a hard time discriminating between Ak, As, and Ag, and between Mk, Ms, and Mg. At the same time, reliability among KEEP observers was falling, partially as a result of problems with these same discriminations. As a result, teacher code 3 was developed.

Teacher Code 3

The final code used for observations of teacher behavior was teacher code 3. This code is identical to teacher code 2 except that Ak, As, and Ag are collapsed into a single category, A., and Mk, Ms, and Mg are similarly collapsed into M.

Reliability

Student Behavior Code

Because of the sweep technique employed in observations of student behavior, the method of code reliability assessment was necessarily a modification of the actual observation procedure. The differences in time required by two independent observers to spot a child required a method which insured that observations on each student began simultaneously.

The procedure thus involved spotting the first student on the student roster. After both observers had found him, they closed their eyes and one observer counted, "1, 2, 3." On the count of "3," the observers opened their

eyes and independently coded the behavior occurring at that instant. The next student on the roster was then spotted and the procedure was repeated.

Depending on time constraints (the procedure was slow), from one to three observations on each student were coded in this manner. After observations were completed, reliability was calculated. The formula used was:

$$\frac{\text{number of identical observations}}{\text{number of total observations}}$$

In this case, only "on-task," "off-task," and "disruptive" observations were entered in the above formula; observations of students "out of field" were not included. The total number of observations refers to the number recorded by each observer, not to their sums.

There was one constant observer, LK, with whom reliability was obtained. Her observations were compared with those of EA, KB, and AM (AM was not used for student behavior data collection). The code was relatively simple and, with a few minutes training, reliabilities of .90 or greater were achieved on the first try.

Later in the year, two additional observers of student behavior, LW and SO, were trained. The former was considered trained in late February when she achieved reliabilities above .85 with LK. Reliabilities achieved in the training of SO are presented in Table 1. Finally, a first semester check on EA-LK reliability and year-end check on the intercode reliabilities of EA, LK, SO, and LW are presented in Table 2.

The data presented in Tables 1 and 2 include not only an "over-all" reliability, calculated as above, but also a series of "category" reliabilities for each category, "on-task," "off-task," and "disruptive." For category reliabilities, the number of identical observations of that category was divided by the total number of observations of that category by the person who recorded

TABLE 1

Student Behavior Code Reliability: Training of SO

DATE	OBSERVERS	I, T, R[a]	RELIABILITY			
			On-Task	Off-Task	Disruptive	Overall
11-16-73	EA	I	81	23	5	109
	SO	T	94	41	17	140
		R	.86	.56	.29	.78
11-19-73	EA	I	83	19	4	106
	SO	T	92	31	10	125
		R	.90	.61	.40	.85
11-30-73	EA	I	94	19	0	113
	SO	T	100	25	2	125
		R	.94	.76	.00	.9[illegible]

[a]I = number of identical observations, T = total number of observations, R = reliability

TABLE 2

Student Behavior Code Reliability; Post-Training Checks

DATE	OBSERVERS	I, T, R[a]	RELIABILITY			
			On-task	Off-task	Disruptive	Overall
11-16-73	EA	I	29	5	2	36
	LK	T	33	9	5	43
		R	.88	.56	.40	.84
6-3-74	EA	I	45	33	12	90
	LK	T	56	50	21	110
		R	.80	.66	.57	.82
6-3-74	EA	I	46	31	17	94
	SO	T	56	40	21	110
		R	.82	.78	.81	.85
6-3-74	EA	I	51	30	10	91
	LW	T	55	39	20	107
		R	.93	.77	.50	.85
6-3-74	LK	I	43	35	12	90
	SO	T	51	50	19	110
		R	.84	.70	.63	.82
6-3-74	LK	I	44	32	9	85
	LW	T	55	48	14	107
		R	.80	.67	.63	.79
6-3-74	SO	I	47	30	11	89
	LW	T	55	39	18	107
		R	.85	.77	.61	.82
6-4-74	EA	I	60	22	6	88
	LK	T	69	37	15	110
		R	.87	.59	.40	.80
6-4-74	EA	I	63	20	2	85
	SO	T	70	33	15	110
		R	.90	.61	.13	.77
6-4-74	EA	I	55	21	3	79
	LW	T	69	44	15	110
		R	.80	.48	.20	.72
6-4-74	LK	I	62	28	2	92
	SO	T	70	37	7	110
		R	.89	.76	.29	.84
6-4-74	LK	I	54	30	2	86
	LW	T	64	44	7	110
		R	.84	.68	.29	.78
6-4-74	SO	I	58	28	0	86
	LW	T	70	44	5	110
		R	.83	.64	.00	.78

[a] I = number of identical observations, T = total number of observations, R = reliability

the most instances of it. The sum of the totals of each category is thus larger than the "overall" total, except in the case of 100% reliability. Consequently, the percentages of each category are usually lower than the overall percentages.

Table 3 gives a summary of reliability on the student behavior code. This table is generated from the data in Tables 1 and 2. For each pair of observers, the sums of identical and total observations from each behavior category are entered in the "I" and "T" rows of Table 3. A weighted mean reliability is then computed as:

$$\frac{\text{sum of identical observations}}{\text{sum of total observations}}$$

For example, from Table 2, it can be seen that EA and LK participated in three reliability sessions (11-16-73, 6-3-74, and 6-4-74). For on-task behavior, the numbers of identical observations were 29, 45, and 60, while totals for on-task were 33, 56, and 69 for the three sessions. Thus, in Table 3 the entry in the on-task column for the "I" row of EA-LK reliability is 29+45+60 = 134, while for the "T" row, the entry is 33+56+69 = 158. The resulting reliability is 134/158 = .85. The grand total in Table 3 is generated similarly, with the number of identical observations defined as the sum of the identical observations for each of the pairs listed in Table 3, etc.

Teacher Behavior Code

Reliability on the teacher behavior code was assessed continuously in minute intervals for 10-15 minutes. Equipped with a stop watch, one observer would signal the start of an observation period. Observers would independently code the behavior of the teacher in question for one minute, recording it on the top line of their recording sheet. After a minute, the second interval was called out by the time-keeper and observers recorded behavior for the next

TABLE 3

Student Behavior Code Reliability; Summary

OBSERVERS	NUMBER OF TRIALS	I, T, R[a]	RELIABILITY			
			On-task	Off-task	Disruptive	Overall
EA LK	3	I T R	134 158 .85	60 96 .62	20 41 .49	214 263 .81
EA SO	5	I T R	367 412 .89	112 170 .66	28 65 .43	507 610 .83
EA LW	2	I T R	106 124 .85	51 83 .61	13 35 .37	170 217 .78
LK SO	2	I T R	105 121 .87	63 87 .72	14 26 .54	182 220 .83
LK LW	2	I T R	98 119 .82	62 92 .67	11 21 .53	171 217 .79
SO LW	2	I T R	105 125 .84	58 83 .70	11 23 .48	174 217 .80
Grand Total	16	I T R	915 1059 .86	406 611 .66	97 211 .46	1418 1744 .81

[a] I = number of identical observations, T = total number of observations, R = reliability

minute on the second line of their sheets, and so on until the end of the period.

Reliability was calculated as the number of identical observations/total observations. Each instance of the same behavior recorded in the same minute was considered an identical observation. Since there was, theoretically, no limit to the number of such behaviors recorded in a given observation session, and since observers often recorded a different total number of behaviors, the "total" for reliability purposes was the number of behaviors recorded by the observer who recorded the most. This reliability figure is an "overall" rather than "category" reliability in the same sense mentioned in the preceding section.

Observers for teacher code 1 were LK and AM. The code was considered developed when reliabilities of .85 or higher were achieved consistently. Since initial development preceded the arrival of the wireless microphone, audio fidelity was poor and most of the training and code development was accomplished with videotapes of teachers in mini-sessions. Several sessions were also run in the classroom with the less mobile teachers.

Work soon began on teacher code 2, again with LK and AM as the initial observers. Mutual training and code development with videotapes and in-classroom observations were finally augmented by sessions conducted from the observation deck when the wireless microphone arrived. Reliabilities in the .80 - .90 range were achieved when observations with teacher code 2 were begun. Spot checks on LK-AM reliability of teacher code 2, conducted during the fall semester are given in Table 4.

TABLE 4

Teacher Code 2 Reliability: Fall Semester

Date	Event
10-5-73	First day of teacher code 2
10-12-73	LK-AM reliability = .82
10-19-73	LK-AM reliability = .84
11-5-73	LK-AM reliability = .91

In January of 1974, EA was trained on teacher code 2. Training was considered accomplished when she achieved .85 or better reliability with both AM and LK.

Reliability figures collected at the beginning of the second semester fell substantially from previous levels. Reliability between LK and AM, taken on 1-15-74 and 1-16-74, was .50 and .65, respectively. Reliability between EA and AM on 1-18-74 was .71. These low reliabilities, in addition to the difficulties in achieving reliability with VISTA volunteers, led to teacher code 3.

Teacher Code 3

It was noted that collapsing "Ak," "As," and "Ag" into a single category "A" and "Mk," "Ms," and "Mg" into "M" (i.e. changing from teacher code 2 to teacher code 3) would substantially raise reliability. The above figures for January reliability are increased as follows: .50 becomes .72, .65 becomes .81, and .71 becomes .76, when the "A" and "M" categories are collapsed as in teacher code 3.

Reliabilities on teacher code 3, given in the following tables, include an "overall" reliability, calculated like that in teacher codes 1 and 2, and a

"category" reliability. The category reliability is the number of identical entries divided by the total number of entries for each coded category of teacher behavior. For coded categories to be identical, the same code must have been entered in the same minute by the two independent observers. The denominator of the fraction is the total number of recordings of the behavior in question within the entire observation session. When observers recorded different totals, the higher number was chosen. As in the case of category reliability of the student behavior code, the sum of the totals across all categories is greater than the total used in computing overall reliability, and category reliabilities are, consequently lower.

Table 5 gives both the category and overall reliabilities of EA, LK, and AM on teacher code 3. The January reliability figures, mentioned above, are also included and are entered as if the code used had been code 3. Table 6 shows reliabilities achieved during the training of a fourth observer, LW (as well as year-end reliability checks). Finally, Table 7 shows reliabilities achieved in the training of KB and SO, two additional observers who were trained as possible back-up observers, but who were never used to collect data on teacher behaviors. In all of these tables, an "x" in the place of a reliability figure indicates that neither observer recorded any instance of the behavior in question in that observation session. It should be noted that such "agreements on nonoccurrence" are <u>not</u> counted as identical observations in any reliability computation.

A summary of reliabilities on teacher code 3, shown in Table 8, is given in a similar manner to the student behavior code reliability of Table 3. Summary reliabilities of pairs of the observers who participated in teacher behavior data collection, either at KEEP or control schools, (EA, LK, AM, and LW), are taken from Tables 5 and 6. In Table 8, an entry in the "I" row for

TABLE 5

Teacher Code 3 Reliability: Observers EA, LK, and AM

DATE	OBSERVERS	I, T, R[a]	RELIABILITY																
			A	M	D	S	NE	SE	T+	P+	H+	B+	T-	P-	H-	B-	B-	O	R	Overall
1-15-74[b]	LK	I	22	2	0							5								29
	AM	T	31	4	1							9								40
		R	.69	.50	.00	x	x	x	x	x	x	.56	x	x	x	x	x	x	x	.72
1-16-74[b]	LK	I	47	0					0			4						0		51
	AM	T	55	1					1			5						1		63
		R	.81	.00	x	x	x	x	.00	x	x	.5[illegible]	x	x	x	x	x	.00	x	.81
1-18-74[b]	EA	I	16	4		4				6		1				1			0	32
	AM	T	19	4		7				8		3				1			2	42
		R	.84	1.00	x	.57	x	x	x	.75	x	.33	x	x	x	1.00	x	x	.00	.76
3-18-74	EA	I	28	13					27			5			4					77
	LK	T	32	18					38			7			8					92
		R	.88	.72	x	x	x	x	.71	x	x	.71	x	x	.50	x	x	x	x	.84
3-18-74	EA	I	25	10					29	1		4			2					71
	AM	T	31	18					38	5		6			6					92
		R	.81	.56	x	x	x	x	.76	.20	x	.67	x	x	.33	x	x	x	x	.77
3-18-74	LK	I	24	12					27	0		5			5					73
	AM	T	30	16					34	5		7			9					88
		R	.80	.75	x	x	x	x	.79	.00	x	.71	x	x	.56	x	x	x	x	.83
6- 3-74	EA	I	38	11					24	2										75
	AM	T	42	15					25	3										83
		R	.90	.73	x	x	x	x	.96	.67	x	x	x	x	x	x	x	x	x	.90
6- 3-74	EA	I	38	13					24	2										77
	LK	T	40	15					25	3										82
		R	.95	.87	x	x	x	x	.96	.67	x	x	x	x	x	x	x	x	x	.94
6- 3-74	LK	I	40	13					23	3										79
	AM	T	42	15					24	3										83
		R	.95	.87	x	x	x	x	.96	1.00	x	x	x	x	x	x	x	x	x	.95

[a] I = number of identical observations, T= total number of observations, R = reliability

[b] Teacher code 2 acutually used but scores adjusted to reflect reliabilities as if teacher code 3 were used

DATE	OBSERVERS	I, T, R[a]	RELIABILITY																
			A	M	D	S	NE	SE	T+	P+	H+	B+	T-	P-	H-	B-	B-	O	R	Overall
2-21-74	AM LW	I T R	5 10 .50	2 4 .50	2 3 .67	1 3 .33	x	x	19 24 .79	0 1 .00	x	3 7 .43	x	x	x	0 1 .00	x	0 1 .00	2 2 1.00	34 46 .74
4- 1-74	EA LW	I T R	8 15 .53	3 5 .60	x	x	0 1 .00	x	115 132 .87	x	x	0 1 .00	x	x	0 1 .00	1 3 .33	x	x	x	127 148 .86
4- 2-74	AM LW	I T R	6 9 .67	3 5 .60	x	0 1 .00	x	x	1 1 1.00	x	x	3 5 .60	x	x	1 1 1.00	x	x	0 1 .00	x	14 20 .70
4- 2-74	AM LW	I T R	29 39 .74	13 18 .72	x	x	x	x	25 33 .76	3 6 .50	x	x	x	x	x	x	x	x	x	70 90 .78
4- 2-74	AM LW	I T R	37 44 .84	7 11 .64	x	x	x	x	5 6 .83	x	x	0 1 .00	x	x	1 2 .50	x	x	x	x	50 61 .82
4- 3-74	AM LW	I T R	24 34 .71	16 24 .67	x	x	x	x	17 21 .81	x	x	1 1 1.00	x	x	x	1 1 1.00	x	x	x	59 69 .855
4- 3-74	AM LW	I T R	10 13 .80	12 14 .86	x	x	x	x	8 12 .75	x	x	4 5 .80	x	x	x	2 3 .67	x	x	0 1 .00	36 44 .82
4- 3-74	AM LW	I T R	21 28 .75	4 5 .80	x	x	x	x	35 35 1.00	2 2 1.00	x	x	x	x	x	x	x	x	x	62 68 .92
4- 4-74	AM LW	I T R	18 20 .90	x	x	x	x	x	8 9 .89	x	x	2 2 1.00	x	x	0 1 .00	x	x	x	x	28 31 .90
4-22-74	LK LW	I T R	8 12 .67	13 15 .87	x	2 3 .67	1 1 1.00	x	16 21 .71	x	x	x	x	x	x	1 2 .50	x	1 1 1.00	3 3 1.00	44 54 .81
6- 3-74[b]	EA LW	I T R	38 39 .97	10 14 .71	x	x	x	x	24 25 .96	2 2 1.00	x	x	x	x	x	x	x	x	x	[illegible]
6- 3-74[b]	AM LW	I T R	39 42 .94	11 13 .85	x	x	x	x	23 24 .96	2 3 .67	x	x	x	x	x	x	x	x	x	[illegible]
6- 3-74[b]	LK LW	I T R	39 40 .98	12 15 .80	x	x	x	x	24 24 1.00	2 3 .67	x	x	x	x	x	x	x	x	x	[illegible]

[a] I = number of identical observations, T = total number of observations, R = reliability

[b] Post-training reliability checks

TABLE 7

Teacher Code 3 Reliability: Training of KB and SO[a]

DATE	OBSERVERS	I, T, R[b]	RELIABILITY																
			A	M	D	S	NE	SE	T+	P+	H+	B+	T-	P-	H-	B-	B-	O	R	Overall
2-12-74	AM	I	12	0					15			0								27
	SO	T	16	2					23			1								36
		R	.75	.00	x	x	x	x	.65	x	x	.00	x	x	x	x	x	x	x	.75
2-12-74	KB	I	22						32			0								54
	AM	T	30						60			1								81
		R	.73	x	x	x	x	x	.53	x	x	.00	x	x	x	x	x	x	x	.67
2-12-74	KB	I	13	0					20											33
	SO	T	20	2					28											54
		R	.65	.00	x	x	x	x	.71	x	x	x	x	x	x	x	x	x	x	.61
2-20-74	KB	I	11	3	0	0			28	1		0				1		0	2	46
	SO	T	20	3	1	1			38	3		5				2		1	5	64
		R	.55	1.00	.00	.00	x	x	.74	.33	x	.00	x	x	x	.50	x	.00	.40	.72
2-20-74	AM	I	13	3		1	0		26	3		3				2			2	53
	SO	T	16	3		1	1		34	4		4				2			5	67
		R	.81	1.00	x	1.00	.00	x	.76	.75	x	.75	x	x	x	1.00	x	x	.40	.79
2-20-74	KB	I	14	3	0	0	0		18	1		1				1		0	5	43
	AM	T	20	3	1	1	1		34	3		4				2		1	5	67
		R	.70	1.00	.00	.00	.00	x	.53	.33	x	.25	x	x	x	.50	x	.00	1.00	.64
2-20-74	KB	I	19	0					40											59
	SO	T	27	1					47											71
		R	.70	.00	x	x	x	x	.85	x	x	x	x	x	x	x	x	x	x	.83
2-20-74	AM	I	20	0					53	1					0					74
	SO	T	33	3					65	1					1					101
		R	.61	.00	x	x	x	x	.82	1.00	x	x	x	x	.00	x	x	x	x	.73

TABLE 7 (continued)

Teacher Code 3 Reliability: Training of KB and SO[a]

2-20-74	KB AM	I T R	21 27 .78	1 2 .50	x	x	x	x	43 46 .93	x	x	x	x	x	0 1 .00	x	x	x	x	65 77 .84
2-21-74	KB LW	I T R	5 9 .56	1 3 .33	1 3 .33	0 3 .00	x	x	20 25 .80	0 1 .00	x	3 5 .60	x	x	x	0 1 .00	x	0 2 .00	2 2 1.00	32 48 .67
2-21-74	KB AM	I T R	9 10 .90	3 3 1.00	1 4 .25	1 2 .50	x	x	40 47 .85	x	x	3 5 .60	x	x	x	0 2 .00	x	0 1 .00	2 2 1.00	59 69 .86
2-21-74	EA KB	I T R	14 19 .74	6 11 .55	x	x	x	x	15 18 .83	x	x	1 2 .50	x	x	2 6 .33	0 1 .00	x	3 3 1.00	1 1 1.00	42 59 .71
2-22-74	KB SO	I T R	2 5 .40	2 5 .67	0 1 .00	x	x	x	x	x	x	x	x	x	1 1 1.00	x	x	x	x	5 9 .56
2-22-74	KB AM	I T R	3 3 1.00	3 4 .75	x	x	x	x	x	x	x	x	x	x	1 1 1.00	x	x	x	x	7 8 .82
2-22-74	AM SO	I T R	2 5 .40	2 4 .50	0 1 .00	x	x	x	2 2 1.00	x	x	x	x	x	1 1 1.00	x	x	x	x	7 10 .70
2-28-74	AM SO	I T R	26 36 .72	2 4 .50	x	x	x	x	27 33 .73	x	1 1 1.00	8 12 .75	x	x	x	x	x	x	x	64 84 .76

[a] KB and SO were trained as possible back-up observers, but they participated in no data collection.

[b] I = number of identical observations, T = total number of observations, R = reliability

TABLE 8

Teacher Code 3 Reliability: Summary of Observers EA, LK, AM, and LW

OBSERVERS	NUMBER OF TRIALS	I, T, R[a]	RELIABILITY																
			A	M	D	S	NE	SE	T+	P+	H+	B+	T-	P-	H-	B-	(B-)	O	R	Overall
EA	2	I	66	26					51	2		5			4					154
LK		T	72	33					63	3		7			8					174
		R	.92	.79	x	x	x	x	.81	.67	x	.71	x	x	.50	x	x	x	x	.89
EA	3	I	79	25		4			53	9		5			2	1			0	178
AM		T	92	37		7			63	16		9			6	1			2	217
		R	.86	.68	x	.57	x	x	.84	.56	x	.56	x	x	.33	1.00	x	x	.00	.82
EA	2	I	46	13			0		139	2		0			0	1				201
LW		T	54	19			1		157	2		1			1	3				228
		R	.85	.68	x	x	.00	x	.89	1.00	x	.00	x	x	.00	.33	x	x	x	.88
LK	4	I	133	27	0				50	3		14			5			0		232
AM		T	158	36	1				59	8		23			9			1		274
		R	.84	.75	.00	x	x	x	.85	.38	x	.61	x	x	.56	x	x	.00	x	.85
LK	2	I	47	25		2	1		39	2						1		1	3	121
LW		T	52	30		3	1		45	3						2		1	3	136
		R	.90	.83	x	.67	1.00	x	.87	.67	x	x	x	x	x	.50	x	1.00	1.00	.89
AM	9	I	189	68	2	1			141	7		13			2	3		0	2	428
LW		T	239	94	3	4			165	12		21			4	5		2	3	512
		R	.79	.72	.67	.25	x	x	.85	.58	x	.62	x	x	.50	.60	x	.00	.67	.84
GRAND	22	I	560	184	2	7	1		473	25		37			13	6		1	5	1314
TOTAL		T	667	249	4	14	2		552	44		61			28	11		4	8	1541
		R	.84	.74	.50	.50	.50	x	.86	.57	x	.61	x	x	.46	.55	x	.25	.62	.85

[a] I = number of identical observations, T = total number of observations, R = reliability

a given pair of observers and for a given behavior category is the sum of corresponding entries in the "I" row from Tables 5 and 6; an entry in the "T" row is derived similarly and the reliability is the ratio I/T. The grand total is obtained in the same way with "I" and "T" summed over all pairs above in Table 8.

It might be noted that in three of the 22 sessions used in computing the grand total, teacher code 2 was used but data were entered as if they had been teacher code 3. Reliability over sessions in which only teacher code 3 was used may be obtained by subtracting out these three sessions. The resulting changes in reliabilities are as follows: D = .67, S = .43, P+ = .53, B+ = .65, B- = .50, O = .33, and R = .83. All other category reliabilities in the grand total are changed by less than .01 when these three sessions are omitted.

Frequency of behavior and reliability

Category reliabilities seem related to the frequency of occurrence of the behavior in question--the more frequent the behavior, the better the reliability. One measure of frequency is the total number of occurrence (T) noted by the observers. This total, the larger of the totals for both observers, is the denominator of the reliability fraction. It is clear that if observations were recorded randomly, the total would be expected to be negatively related to reliability. There is, however, a strong positive relationship.

Across categories, the relationship can be assessed from the grand totals from the student code and teacher code 3 summary tables (Tables 3 and 8). The correlation between totals for the student code and reliability ("T" and "R" of "Grand Total" in Table 3 across behavior categories, not overall) is .999. The corresponding correlation for teacher code 3 ("T" and "R" rows for "Grand Total in Table 8 across behavior categories and excluding those categories noted by "x" is .823. These relationships, although quite strong, give only

partial evidence of a relationship between frequency and reliability because of the cross-category compounding.

Within categories, however, the relationship is still found. Table 9, demonstrating this relationship, is generated as follows. For each category of both the student behavior code and teacher code 3 (from Tables 1, 2, 5, and 6), the mean "T" ($\bar{T}$) was found. Since reliability is undefined (0/0) when behavior of a given category is not observed by either observer, the mean is taken across only those sessions for which at least one instance of the behavior in question was recorded by at least one of the observers. In other words, $\bar{T}$ was computed only across those sessions where the reliability for the behavioral category in question is not denoted by "x". Thus, for example, from Tables 5 and 6, there were 22 sessions in which at least one observer noted at least one instance of "A" behavior, but only 21 qualifying sessions for "M" behavior (neither observer recorded an "M" on 4-4-74 in Table 6). The mean for "A" is taken over 22 sessions, while for "M" the mean is taken over 21 sessions.

Reliability sessions in which "T" is higher than $\bar{T}$ for a given category are classified as high frequency for that category; reliability sessions in which "T" is lower than $\bar{T}$ are, conversely, classified as low frequency. Sums of "I" and "T" for high and low frequency sessions, taken across all pairs of observers in Tables 1, 2, 5, and 6, are used to generate high and low frequency category reliabilities. (In the case of the "O" and "R" categories, some sessions had a total exactly equal to $\bar{T}$. Here, half of the "I" and "T" for such sessions was summed with the high frequency and half with the low frequency sums).

From the difference row in Table 9, it can be seen that of the 15 categories considered, 11 are higher in sessions in which at least one observer noted a high frequency of that behavior. High frequencies are associated with

TABLE 9

Category Reliabilities: High vs. Low Frequencies

CATEGORY[a]		STUDENT BEHAVIOR CODE			TEACHER CODE 3											
		On Task	Off Task	Disruptive	A	M	D	S	NE	T+	P+	B+	H-	B-	O[b]	R[b]
M_T		66.2	38.2	13.2	30.3	11.9	2.0	3.5	2.0	26.3	3.4	4.7	4.0	1.8	1.8	.1
High Frequency	N[c]	8	10	10	13	13	1	1	1	6	4	8	3	3	2	2
	R[d]	.88	.67	.50	.87	.76	.67	.57	.67	.83	.42	.62	.48	.50	.25	.1
Low Frequency	N[c]	8	6	6	9	8	1	3	1	15	9	5	4	3	2	2
	R[d]	.84	.66	.28	.74	.64	.00	.43	.00	.89	.75	.50	.40	.67	.25	.33
Difference[e]		+.04	+.01	+.22	+.13	+.12	+.67	+.14	+.67	-.06	-.33	+.12	+.08	-.17	.00	-.23

[a] Categories never observed in reliability sessions (SE, H+, T-, P- and B-) are omitted.

[b] In the "O" and "R" categories, sessions where the total number of observations = M_T are delegated half to high frequency and half to low frequency.

[c] Number of sessions with category totals indicating high vs. low frequency.

[d] Reliability = sum of identical observations/ sum of total number of observations in the N sessions.

[e] High frequency reliability - low frequency reliability.

lower reliabilities in only three categories (T+, B-, and P+), while there is no difference in a final category (0).

The mean difference per category, for the student behavior code is +.09, and mean difference, per category, for teacher code 3 is +.15. In other words, for a category chosen at random from the student code, a "grand total" reliability, summed over sessions in which the behavior is relatively frequent for that type of behavior, the reliability is expected to be about .09 higher than would be a comparable reliability summed over low frequency sessions. Similarly, a category chosen at random from teacher code 3 will be expected to be about .15 higher when high frequency sessions are considered.

APPENDIX A

Discrepancies From Prescribed Order of Student Observation

Occasionally, the problem of the lack of synchronicity in the starting of various classes resulted in observations of students in a sequence different from that intended. For example, in the spring semester, morning classes began with Reading Groups, Group one then two for one teacher and Group four then five for another teacher. During this time, students not in reading groups were in learning-centers. Some observations were scheduled for the time of the second reading session in the morning (Reading Groups 2 and 5 in reading class and the rest of the students in centers). If one teacher spent less time with Reading Group 1 than the other spent with Reading Group 4, then Reading Group 2 would begin before Reading Group 5. The normal procedure would be to wait until Reading Group 5 started before initiating observations. Occasionally, however, Reading Group 4 would last well into the session with Reading Group 2.

In such a case, rather than missing an entire observation session, observations were begun before Reading Group 5 begun. But, some students were in the "wrong class" (e.g. in the above example, students in Reading Group 4 who should be in learning-centers at the time of observation were with the reading teacher, and students in Reading Group 5 who should be in reading were still in learning-centers). Such students were skipped in the observations. But, if during the observation, the reading class switched so that the students were in the "right" class, they were observed. The result, in such cases, was a discrepancy from the prescribed order of observations.